DOCUMENT RESUME

ED 185 979 IR 008 308

AUTHOR Nielsen, Per
TITLE European Social Science Data Infrastructure: Potential Links to the Library Sector.
PUB DATE Aug 79
NOTE 15p.; Paper presented at the Conference of the International Federation of Library Associations and Institutions (45th, Copenhagen, Denmark, August 27-September 1, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Archives; Data Bases; Documentation; *Information Dissemination; *International Organizations; Libraries; Library Cooperation; *Social Sciences
IDENTIFIERS *Europe

ABSTRACT

This presentation on the European social science data organizations and such international cooperation in federations/associations as CESSDA (Committee of European Social Science Data Archives), IFDO (International Federation of Data Organizations for the Social Sciences), and IASSIST (International Association for Social Science Information Service and Technology), is extended into an outline of the data resources available and a discussion of the output relevant to traditional libraries. A platform for data organization-library cooperation in the field of documentation is proposed, based in part on the experiences gained in the North American branch of IASSIST. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

IFLA General Council 1979

Division: SPECIAL LIBRARIES

Section: SOCIAL SCIENCS LIBRARIES

ED185979

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EUROPEAN SOCIAL SCIENCE DATA INFRASTRUCTURE:

POTENTIAL LINKS TO THE LIBRARY SECTOR.

BY

Per NIELSEN, Archive Director
Danish Data Archives
Odense University
Niels Bohrs Allé 25
DK-5230 Odense M, Denmark

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carol Kuceri Henry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR 008 308

July 1979 (M-1278)

European Social Science Data Infrastructure:

Potential Links to the Library Sector.

0. Abstract

A presentation of the European social science data organizations and the international co-operation in federations/associations like CESSDA, IFDO, and IASSIST is extended into an outline of the data resources available and a discussion of the output relevant to traditional libraries. A platform for data organization - library co-operation in the field of documentation is proposed, in part based on the experiences gained in the North American branch of IASSIST.

1. Institution-Building in Europe

In the pre-information era, problems related to preservation, storage and access to social science material were successfully solved by the library and the document archive:

> Before the arrival of the computer, these giant data-collection agencies cooperated without great difficulty with the data-storage institutions: the tables and the analyses published by the statistical bureaus were stored in libraries and the original data sheets (census sheets, register protocols, and the like) were with some regularity transferred to the established archives.[1]

In the article on Data Services in Western Europe quoted above, Stein Rokkan claims that "the inertia of the traditional institutions" in adjusting to the storage and display demands after the computer revolution created an unsatisfied demand for mass data in computerized form; the Data Organizations (DOs)[2] tried to bridge the gap between production and distribution vis-a-vis the social science community.

Data organizations with national coverage were set up in several European countries during the sixties and early seventies; a framework for cooperation among the more established DOs in West Europe was provided by the Committee of European Social Science Data Archives (CESSDA), set up in Amsterdam in 1976. The founding (and present) members of CESSDA are:

Archivio Dati e Programmi per le Scienze Sociali, Milan (ADPSS)

Belgian Archives for the Social Sciences, Louvain-la-Neuve (BASS)

Danish Data Archives, Odense (DDA)

Norwegian Social Science Data Services, Bergen (NSD)

Social Science Research Council Survey Archive, Essex

Steinmetzarchief, Amsterdam

Zentralarchiv für empirische Sozialforschung, Cologne (ZA)

CESSDA was founded as an informal body. Participation is open for all established Social Science Data Archives and Data Services in West Europe. General meetings are held once a year. The Committee has a rotating office, currently the Zentralarchiv, which is that of the secretary of the next annual conference. All interarchival activities are to be reported to the committee and should be published in the archive newsletters. Furthermore emphasis is on exchange of information on the archive holdings and regulations concerning the exchange of data. All members agreed to sponsor European working parties.

Facing the need to have an international body as a reference point for the DOs CESSDA called a meeting in Louvain-la-Neuve. The International Federation of Data Organizations for the Social Sciences (IFDO) was founded there in May 1977 - the founding members being the seven CESSDA-archives and four North American Data Organizations. All organizations which have the

capacity and expertise to promote the exchange of data and technologies among data organizations can apply for IFDO membership. The general assembly explicitly acknowledged the activities of IASSIST (see below) and incorporated in its articles of governance (article IX) to establish close working relations with IASSIST. Before outlining the objectives of IFDO, the latter association is introduced.

2. The professional membership organization IASSIST

The International Association for Social Science Information Service and Technology (IASSIST) was founded at the World Congress of Sociology 1974 in Toronto. The objectives of IASSIST were defined as follows:

2.1. To encourage and support the establishment at local and national levels of information centers for data base reference, maintenance, and dissemination.

2.2. To foster international dissemination and exchange of information on significant developments in information centers for statistical and textual machine-readable data bases.

2.3. To coordinate on an international level programs, projects, and general procedural efforts which provide an international forum for the discussion of problems relating to information centers.

2.4. To promote the development of professional standards and encourage the establishment or training courses for data center personnel.

In particular the North American IASSIST Groups were very active in the past years and have come up with substantial suggestions for improving data services. Most of the results were published in the IASSIST Newsletters.

In Europe, on the other hand, it has proven difficult to bridge the gap between the academic DOs (from whose personnel most European IASSIST members were initially recruited) and other documentation areas. To the library sector and to the document archives the links of communication have been very weak. However, with the creation of national and international information systems, the understanding of and sensitivity towards documentation requirements is emerging in tradftional archives and in libraries. Part of this change in attitudes may be explained by increasing familiarity with data processing techniques; the application of a computer for data analysis and documentation is not so cryptic any more.

Even though the gap at the technical level of mutual understanding is closing, there remains a lacking insight into the substantive and methodological problems of data documentation. This is particulary true in the areas of data preparation procedures and background documentation. Again, there is a good ground for the expectation that after the linking and interconnecting of the "DO-type" data documentation (see section 4 below) with related activities in the "traditional" documentalist sectors the gap will be overcome rather quickly.

Due to the limited personnel resources in Western Europe IASSIST activities have to be closely coordinated with the activities in CESSDA and IFDO. We shall briefly introduce the IFDO plans for activity and later return to a more detailed description of areas where the librarians and the DO-personnel may cooperate.

3. The IFDO activities

The IFDO activities planned in 1977 reflected a transfer of tasks from the CESSDA-group to this international organization; similarly, many detail tasks have been solved in a cooperative effort with IASSIST. The following outline of activities reflects with some generality the con-

3.1. Training seminars and technology transfer:

Most of the IFDO members have accumulated particular knowledge in social science data processing for archival purposes as well as for data analysis. This experience could be transferred to sites which are less advanced, technologically and institutionally. One project will be a "Seminar for personnel in technologically weak countries".

Another project will concentrate on the promotion of archival holdings and more advanced analysis strategies. A work group has been set up to compare experiences with various "Workbooks, training seminars and related teaching aids" and to come up with suggestions for further development.[4]

3.2. Archive operations and tools

Even though the major archives are facing by and large the same data manipulation problems they have developed a variety of procedures, which are only roughly comparable. An ever lasting problem tends to be that there is no satisfaction with existing programs for handling multiple punch data. While the development and standardization of program routines will take more time, priority was given to defining an "International Standard Data Quality Description". This should make the various existing quality categories and status classes comparable.

Data documentation could also benefit from more inter-archival compatibility. On the data file level an effort had been made to develop the Standard Study Description Scheme (see section 4 below). An attempt will be made to implement the agreed upon formats after a testing phase - which may lead to further modifications.[5]

In a joint effort the Danish Data Archives and the Zentralarchiv have been working towards a Standard Codebook Format. The list structuring program used is OZPRINT. The character OZ indicate compatibility with OSIRIS codebook formats and Z.A.R. retrieval formats. A joint IASSIST/IFDO project was suggested by the general assembly to examine codebook documentation and to develop archive standards. First steps have been taken in that direction.

3.3. Promotion of data usage

Training seminars and teaching packages will bring the potential user closer to the archive holdings. Much can be done, however, to make the archives' information potential more attractive. An initial step could be to compile collections of high standard studies. By checking their contents against the holdings of other archives the chance of constructing comparative studies ex post is given.

3.4. Recent topics in data services

On the initiative of the Norwegian Social Science Data Services the establishment of a joint data bank for thematic mapping of regional variations in Europe has been discussed. The NSD has established a facility for computer mapping of variations not only across the communes of Norway but even across the "level one" regions of Western Europe and - in a Joint Nordic Project - across the county (amt / fylke/ län/syslu) level in the Nordic countries.[6] A CESSDA joint project is underway here, since the BASS and the ADPSS have agreed to join in the development of the data system and in the production of the first computer atlases.[7]

Another topic of particular concern is Privacy legislation and research". The IASSIST Newsletter has discussed the issue of confidentiality. CESSDA and IFDO co-sponsored the first international conference on "Emerging Data Protection and Social Sciences' Need for Access to Data" (Cologne, August 1978). This conference was intending to report on national developments and propose solutions to the problems inherent with the data legislation in most western countries.[8]

4. "DO-type" data and documentation

The most obvious field where cooperation between the DOs and the libraries is urgently needed is that of data documentation. Based on an IASSIST workshop in Copenhagen (June 1977)[9] the list of relevant documents and facilities concerning data stored at DOs may be compiled as outlined in eight subsections

below; before turning to the documentation, however, a couple of paragraphs on the data stored with the DOs are due.

Historically, political scientists took the lead in the "data archive movement", the typical data types being opinion polls and other mass surveys. During the first moving decade, however, quantitatively oriented social scientists of almost any discipline entered the secondary analysis arena along with policy makers and planners in public agencies and private organizations. As data producers and secondary analysts (in relation to the DO: donors resp. users) are the same groups, the substantive fields covered and the data types represented in archive holdings broadened considerably.

Even so, the bulk of the datasets in the DOs are rectangular files (i.e. simply structured data matrices with a given set of information items (variables) for all surveyed units); furthermore, the mass survey is prevailing in the holdings of most of the DOs. It is only recently that a considerable amount of energy is being devoted towards hierarchical and relational data structures.

Consequently, the list of documents/facilities below is particularly relevant to data originating in survey research and organized in a rectangular matrix.

4.1. Library information:

Information on a data-file should be produced in a card catalogue form to be entered into the library system. Till this day, the inertia of the library systems in Europe has been so strong that data files are not generally considered for reference. In the U.S. the communication between DO-personnel and traditional librarians is better: Coordinated by Sue A. Dodd and following recommendations from the ALA Catalog Code Revision Committee's Subcommittee on Rules for Cataloging Machine-Readable Data Files, the Classification Action Group of IASSIST in North America has set standards that should be adopted also in Europe.[9]

4.2. Archive information / study abstract

Each data service organization has its own way of presenting its holdings (Inventory). It would be very useful if other data holders (e.g. statistical bureaus, machine readable divisions of traditional archives, research institutions) would publish guides/inventories to their holdings. Abstracting of data file contents (cp. the immense resources invested in bibliographic abstracting these years) is hardly seen in Europe outside the realm of the DOs organized in CESSDA.

4.3. Study Description:

The Study Description is the locus for all information on a data file necessary to secure correct interpretation of analysis results, i.e. a complete description of the research project or administrative process in which the data was originally collected and processed. This is probably the area where the sins of omission on the part of the data producer are most outstanding: Not only is it hard to find a full description when a DO acquires a dataset, but, we still see reports on quantitative research results containing incomplete technical report sections. Under the auspices of the former Standing Committee for Social Science Data of the ISSC a Standard Study Description Scheme was developed in 1974, and this scheme has been tested further by several DOs.[11] The Standard Study Description Scheme has been designed in a flexible way to allow for a multi-purpose use: Besides containing all background information necessary for a reliable secondary analysis of the file described, the information can be automatically subsetted for abstracting purposes; furthermore, the sum of Study Descriptions in a DO can be used for mapping and methodological research as well as for information retrieval and intra-archival logging purposes; finally, the Study Description is used for inter-institutional exchange of information on data holdings. It should be repeated, however, that the standard study description in its present version is geared to survey files, primarily.

4.4. List of variables:

Each data organization should produce a list of variables for every file in their holdings; this list of variables should be printed as well as machine-readable. Like the data abstracts may be automatically generated from the study description, this list of variables may be automatically produced from a machine-readable codebook.

4.5. Codebook:

In the Copenhagen IASSIST workshop[9] it was indicated that the ideal codebook should contain 15 major items (the first three items referring to file level, the last 12 to be applied variable by variable where applicable): (1) title of study (file and subfile names); (2) format of the data file; (3) comments at file level, e.g. concerning application of missing data codes, special weighting features, special precautions for use of file; (4) variable identification (number, label, short name, mnemonics);

(5) variable source reference; (6) variable location and length; (7) variable type (alphabetic, alphanumeric, numeric, symbolic); (8) number of decimal places (scale of measurement); (9) source statements/texts/questions/scale description/introductory statements related to responses; (10) answer code values; (11) answer code descriptions; (12) comments originating in field work and coding experience: interviewer instructions and coding instructions; (13) variable contingencies (filter, skip, probe, control); (14) variable consistency (i.e. results of checking of variable contingencies); (15) reference to derived variables.

4.6. Classification/index:

Within the area of classification and indexing (on file and/or variable level) a number of different schemes are available, describing different dimensions. However, the testing of these schemes has not yet reached a level where recommendations can be set forth. For a discussion of variable level retrieval, see the description by Ekkehard Mochmann[12] who has been coordinating the European classification efforts within IASSIST.

4.7. The data matrix:

Even if the data itself can only indirectly be called documentation, it is documenting the results published by the primary investigator; the data matrix is the object that is described in the data documentation. It should be borne in mind that machine-readable data (unlike paper-carried information) can be read only if a correct description of the data carrier and the physical and logical characteristics concerning representation of data on the data carrier is supplied.

4.8. Special publications:

Many data organizations produce publications concerning specific files. Items 2.3 - 2.5 above is often making up several hundred pages concerning one survey. Such data documentation publications should, of course, be entered into the library system. Very often, all data documentation is machine-readable.

5. Proposed DO-library cooperation in Europe

The problem that we are facing has been phrased in the following manner by the afore-mentioned Sue Dodd:[13]

> In the last two decades, private research organizations, government agencies, and foundations have invested heavily in the collection of social science numeric data, contributing to the proliferation of machine-readable data. However, the development of information technology and the ability to produce data have progressed much more rapidly than our capacity to organize, classify, and reference its availability.

In the DOs it is the feeling that we have invested a considerable effort in improvements of the organizing, classifying, and referencing capacity - and that we now need an access to the information dissemination channels of the library sector.

Ideally, the DOs have for each of their datasets:

5.1 The data matrix itself in a clean version and in an easily analysable (by means of existing program packages) format.

5.2. An Extended Study Description including items 4.1 - 4.3 above. This document (10-12 pages) is available to librarians and other information dealers for cataloging, abstracting, and other referencing purposes.

5.3 A Data Documentation Publication, containing the Extended Study Description as well as a list of variables and the Standard Codebook, i.e. items 4.1 - 4.5 above. Sometimes, this "publication" is just a stapled set of xerox copies or computer lists; in other cases, it takes the form of a printed book.

The DO-library cooperation has the well-defined purpose of establishing "automatic" card catalogue and citation/abstract reference production and dissemination concerning data files. In cases where the Data Documentation Publications take the form of printed books, these may be referenced separately - and such books should be available in the libraries.

Some of the preparatory work has been done in IASSIST in North America where the IASSIST membership list counts many librarians. In Europe, librarians would be welcomed in IASSIST which is an association structured in working groups; we need a working group on bibliographic references for numeric social science data files.

The ultimate goal would be to pave the way for social science data files to be included in printed bibliographies, end-of-work references, and indexing and abstracting works such as the Social Science Citation Index[14].

REFERENCES AND NOTES:

1. Stein Rokkan, "Data Services in Western Europe. Reflections on Variations in the Conditions of Academic Institution-Building", American Behavioral Scientist, Vol. 19, No. 4, March/April 1976 (p. 445). This issue of the ABS edited by Richard I. Hofferbert and Jerome M. Clubb, deals with the "data archive movement".

2. The academic data archives (data services, data libraries) will in the following be referred to by the term "Data service organization" or just "data organization". The acronym DO will be used throughout the article.

3. The activity outline, as the whole first part of this paper, draws on an uppublished paper read at the Panel on Documentation of the North American IASSIST Annual Conference on "State of the Art: Perspectives," held at Itaska, Illinois, February 1978:
Ekkehard Mochmann and Per Nielsen, "The Data Documentation Scene from the European Perspective."

4. A Seminar on Data Based Teaching under the auspices of IFDO is scheduled for November 1979 in Waterloo, Canada.

5. A CESSDA Meeting is planned in Luxembourg (September 1979) where a final version of the standard Study Description Scheme should emerge.

6. The Joint Committee of the Nordic SSRCs is sponsoring this project along with other projects proposed by a Subcommittee on Joint Nordic Data Services.

7. An IFDO Seminar concerning ecological data and thematic cartography is planned for the spring of 1980 in Turin, Italy.

8. Resolutions from this conference have been widely circulated; Proceedings (available in a preprint form) are presently being published (Frankfurt: Campus; expected November, 1979).

9. A report on the West European IASSIST Workshop in Copenhagen given in the IASSIST Newsletter vol. 1, no. 4, 1977.

10. For further details, see "Working Manual for Cataloging Machine-Readable Data Files", compiled by Sue A. Dodd, Data Librarian, Social Science Data Library, Institute for Research in Social Science, Univ. of North Carolina, Chapel Hill. (cf. note 13 below.).

11. See "Report on Standardization of Study Description Schemes and Classification of Indicators" and "Study Description Guide & Scheme", both edited by Per Nielsen (Sept. 1974 and April 1975, respectively). Both available from the Danish Data Archives. Since 1974, the ZA and the DDA have been using the Standard Study Description Scheme, and the Steinmetzarchief has been using a modified version thereof. From 1977, also BASS and several North American DOs have tentatively used the scheme.

12. Ekkehard Mochmann, "Information Access at the Data Item Level: Approaches to Indicator Retrieval from Survey Archive Data Bases", SIGSOC Bulletin, Vol. 6, Nos. 2 & 3, Fall/Winter 1974-75.

    Edited by Alice Robbin, this issue of the Bulletin dealt with "The Data Library: Systemic, Structural and Process Problems of Data Access".

13. Sue A. Dodd, "Bibliographic References for Numeric Social Science Data Files: Suggested Guidelines". Journal of the American Society for Information Science, March 1979, p. 77.

14. Ref. 13, p. 82.